

The Ohio Youth Assessment System

Interrater Reliability Study

Edward Latessa, Brian Lovins, and Jenni Lux
University of Cincinnati, Center for Criminal Justice Research
School of Criminal Justice
University of Cincinnati

PURPOSE

This study is an extension of the initial pilot study conducted in 2010. The pilot study was conducted on three courts across Ohio to examine the process in which staff were using the OYAS and to provide feedback regarding interviewing skills and scoring. Hamilton, Franklin, Miami, Belmont, Warren, and Montgomery Counties were invited to participate; of which we observed interviews conducted in Franklin (2), Warren (1), and Belmont (7). The observations were conducted by a CCJR staff person trained in the OYAS. The CCJR staff observed individual interviews conducted by court personnel. Part II of the interrater reliability study was conducted during FY 2011. Part II of the study will examine the interrater reliability of the OYAS Disposition, Residential, and Reentry tools. Interrater reliability measures the level of agreement between raters on a given measure, determining if the tools are scored given the raters understanding of the scoring guide. Specifically, this study will examine the level of agreement between OYAS end users using two statistical techniques—interrater agreement and Krippendorff’s Alpha. The results of this study will be used to inform future trainings and quality assurance efforts.

INTRODUCTION

The Ohio Youth Assessment System is a series of assessments that were developed specifically for Ohio to be used to assist in making decisions regarding youth in the juvenile justice system. The OYAS covers each stage of the juvenile justice system, providing tools for diversion decisions, assisting in detention decisions, disposition, residential intake, and residential reentry. Each tool is designed to measure risk at each stage, offering treatment targets unique to the population being served.

As a fourth generation assessment tool, one of the primary benefits is its ability to measure change over time through the inclusion of dynamic risk factors or criminogenic needs. Although useful in determining dynamic risk, Baird (2009) cites the difficulty in accurately scoring these items, suggesting that interrater agreement typically falls below an acceptable level. To address this concern, Ohio has implemented a comprehensive assessment protocol that includes final certification of all end users. The training protocol includes a 2 day face-to-face training, a written test, and a video scoring test. This process was designed to ensure that staff who will be using the OYAS can achieve a valid and reliable score prior to certification. Now that the OYAS has been in-place for almost two years, it is necessary to gauge the level of reliability the tool has maintained with end users. This study will address whether the OYAS maintains interrater agreement post-certification.

METHODS

Part I of the study was conducted in May 2010. Three counties were selected to participate in Part I of the study. Part II of the interrater reliability study was conducted from April, 2011 to June, 2011. Six counties and ODYS were selected to participate in the study.¹ To conduct an interrater reliability analysis, it is necessary for a minimum of 2 raters to score a youth on the same information. To achieve this, we asked that each staff participating in the study to observe an interview conducted by one of their peers and score the OYAS independently based on the same information. This process ensured that the scores were based on the same information, and that the scores did not reflect the differences in interviewing skills or abilities to obtain collateral information. Once the assessment was completed and the tool scored, the staff was asked to submit the scores to UCCI for analyses. A total of 190 staff participated resulting in 95 youth being assessed. To conduct the analyses, two statistical methods were used. First, interrater agreement was calculated across all the pairs. This provides the average agreement across each item of the tool. Second, Krippendorff's Alpha (Kalpha) was calculated for each pair of raters on their total agreement and then averaged across all pairs. Kalpha is a measure of agreement that provides a more conservative measure of agreement since it controls for the likelihood that two raters would score the same on an item by chance.

FINDINGS

Part I of the study found a consistent use of the tools and scoring materials across the three counties. Staff was generally polite and respectful to the youth and his/her family. The staff that conducted the assessment was knowledgeable about the scoring rules and used the tools provided to score the instrument accurately. They conducted interviews in settings that ranged

¹ The following counties/departments participated in the study: ODYS & Belmont, Franklin, Mahoning, Richland, Lorain, Seneca, and Lake Counties.

from semi-private to private offices. The majority of staff provided an appropriate introduction to the youth regarding the purpose of the interview. Several staff did not share with the youth the purpose of the interview, but instead started right into asking questions.

As for the interview process, there was a range of interviewing skills across the staff. The staff who had previous training in interviewing skills did significantly better interviewing youth than those staff who had not been trained in effective interviewing skills. The interviews ranged from 25 minutes to 3.5 hours. The 3.5 hour interview was integrated into an assessment that was significantly broader than the OYAS. Excluding this assessment, the average time it took staff to conduct the interview was 48 minutes. Although several staff conducted effective interviews, the major limitation to the interviews was the lack of follow-up questions. The assessment is designed to be conducted in a semi-structured process in which staff is highly encouraged to continue probing when information is not clear. There were several times that the staff had to gather more direct information to score the tool, but did not take the opportunity to ask a set of follow-up questions. In addition to the interviewing skills of the staff, CCJR staff also noted that one of the counties had modified the self-report and that very few of the staff used the scoring guide to assist in scoring the tool. Furthermore, two of the counties had modified the interview guide (which is appropriate) but had left out several key questions that needed to be asked to score the tools accurately.

In addition to the observation of the interviews, CCJR also examined the level of agreement between court personnel and CCJR staff scoring. Overall, there was a strong agreement between the county staff and the CCJR staff. Based on the 10 interviews the interclass correlation was .977. Generally correlations above .700 are considered appropriate and demonstrate a strong correlation between scoring items. Individually, all of the interviews

reached a correlation of .712 or higher with 6 of the interviews meeting correlations of .900 or higher. These scores demonstrate a strong correlation between the interviewer and the CCJR staff person's scores.

Based on the results of the Part I study, it was determined that a larger study be conducted to ensure that the results found in the Part I reflected other county's experiences with the OYAS. Part II of the study was conducted slightly different than Part I. As noted in the Method's Section Part II paired non-CCJR staff together to examine the level of agreement between staff in the field. Table 1 provides the interrater agreement for the OYAS Disposition Tool. Again, agreements above .70 are routinely identified as "good." As noted, the overall agreement for the tool was .90 demonstrating high agreement. Individual items ranged from 78 percent to 100 percent suggesting that all items were within an acceptable range. Specifically, 19 of the 32 items on the Disposition Tool were rated with 90 percent or higher agreement, and all but 1 item were rated 80 percent or better. The only item that fell below 80 percent was *Item 5.1 Can Identify Triggers/High Risk Situations* which was rated in agreement 78 percent of the time. Although well above the threshold of 70 percent agreement, the Pro-Social Skills domain (5.1, 5.2, & 5.3) had the lowest percentage of agreement at 83 percent agreement. The overall agreement of the Disposition Tool suggests that raters in the community are scoring the tool with high agreement. In examining the Kappa for the overall disposition sample, the Kappa score was .80, again falling well within an acceptable range of .70 or higher.

Table 1: Percent Agreement OYAS Disposition Tool

1.1 Documented Contact with Juvenile Justice System	94%
1.2 Previous Adjudications	84%
2.1 Family is Important	98%
2.2 Consistently Applies Consequences	83%
2.3 Follows Caregivers Rules	86%
2.4 Follows Through with Consequences	89%
2.5 Contact with Biological Adoptive Parent	97%
2.6 Relationships with Adults	91%
3.1 Friends Fight	84%
3.2 Friends Arrested	80%
3.3 Friends Family Associated with Gang Activity	97%
3.4 Arrested with Friends	91%
3.5 Friends Suspended Expelled from School	92%
3.6 Friends are Important	83%
4.1 Suspended from School Ever	100%
4.2 Suspended from School Last 6 Months	95%
4.3 Expelled Ever	92%
4.4 Relationship with Current School Personnel Employer	92%
5.1 Can Identify Triggers High Risk Situations	78%
5.2 Weighs Pro Cons of a Situation	80%
5.3 Pro Social Decision Making	91%
6.1 Age of Drug Onset	95%
6.2 Used Drugs Recently	97%
6.3 Used Alcohol Recently	98%
6.4 Likely to Quit	92%
6.5 Inflated Self Esteem	86%
6.6 Mental Health Issues	87%
7.1 Pro Criminal Sentiments	92%
7.2 Future Criminal Behavior	89%
7.3 Blames Others	83%
7.4 Attitude Towards Gang	97%
7.5 Self Efficacy	91%
Overall OYAS Score	90%
	Kappa
Overall Kappa Level	0.80

Table 2 provides the results of the OYAS Reentry Tool reliability analysis. As noted, the overall level of agreement was high at 90 percent. The individual OYAS Reentry items ranged from 74 percent agreement to 100 percent agreement, again suggesting that all 41 items on the Reentry tool fall well within an acceptable range of agreement. Specifically, 24 of the 41 items were rated at 90 percent agreement or higher, with 39 of the 41 items above 80 percent. The two items 5.3) Pro-social decision making and 3.9) Adults in the Community are Supportive received the lowest agreement at 76 and 74 percent respectively. Similar to the OYAS disposition tool, the Pro-Social Skills domain received the lowest level of agreement at 79 percent. The overall Kappa was .81, suggesting that even controlling for random responses that the reliability of the tool remains high.

Table 2: Reentry Interrater Reliability

1.1) Documented Contact with Juvenile Justice System	100%
1.2) Attempted and/or Escaped from Residential Facility	95%
1.3) History of Selling Drugs	93%
1.4) Physical Altercation with an Authority Figure	83%
1.5) Weapon Used During a Crime	98%
1.6) Victim Physically Harmed During Offense	86%
1.7) Received Major Sanction while in Residential Care	88%
2.1) Family is Important	95%
2.2) Family Member(s) Arrested	95%
2.3) Parents Use Appropriate Consequences	81%
2.4) Positive Relationships with Person at Planned Residence	93%
3.1) Acquaintances Use Drugs	95%
3.2) Friends Fight	90%
3.3) Friends Use Drugs	90%
3.4) Friends Arrested	93%
3.5) Relationship with Youth on Unit	98%
3.6) Relationship with Staff	98%
3.7) Friends/Family Associated with Gang Activity	88%
3.8) Arrested with Friends	90%
3.9) Adults in the Community are Supportive	74%
4.1) Truant from School	98%

4.2) Expelled Ever	88%
4.3) Effort in School	88%
4.4) Relationship with Current School Personnel/ Employer	98%
5.1) Can Identify Triggers/High Risk Situations	81%
5.2) Weights Pro/Cons of a Situation	81%
5.3) Pro-social Decision Making	76%
5.4) Frustration Tolerance	86%
6.1) Age of Drug Onset	95%
6.2) Others Complained about Drug/Alcohol Use	95%
6.3) Positive Drug Test within Past 6 Months	98%
6.4) Alcohol/Drugs have Caused Problem in Major Life Area	95%
6.5) Used Substances while in Residential Facility	98%
6.6) Inflated Self-Esteem	93%
6.7) Risk Taking Behavior	83%
7.1) Pro-Criminal Sentiments	93%
7.2) Negative Attitudes Towards Supervision	83%
7.3) Attitude Supports Substance Use	86%
7.4) Demonstrates Remorse for Offense	83%
7.5) Demonstrates Empathy Towards Others	88%
7.6) Attitude Towards Gangs	90%
Overall OYAS Score	90%
Kappa	
Kappa Score	0.81

DISCUSSION

Overall, the level of interrater agreement on the OYAS is within acceptable range. Specifically, each tool demonstrated overall agreement above 80 percent and Kappa scores above .70. These scores suggest that the OYAS materials, training process, and certification process produces high rates of interrater agreement, even on items that have been traditionally difficult to obtain agreement.

As expected, the items that had the lowest level of agreement were associated with the Pro-Social Skills Domain. The Pro-Social Skills domain is comprised of three items (Disposition Tool) and four items (Reentry Tool) including whether the youth can identify high risk situations, weigh the pros/cons of a situation, make pro-social decisions, and the level of frustration tolerance. These items can be difficult to score because they ask the assessor to judge the youth's thought processes in making decisions not just the final decision. This is evident in the Dispositional Tool where the level of agreement was 10 percent higher for pro-social decision making (a behavioral indicator) than for identifying high risk situations and weighing the pros/cons of a situation.

RECOMMENDATIONS

Several recommendations can be offered to ensure that the OYAS is conducted effectively:

1. There was some evidence that counties have modified portions of the tool that directly relate to the validity of the tool.
 - a. Ensure that all counties understand what portion of the OYAS can be modified and clarify that the self-report cannot be modified.

2. Staff did not always consult the scoring guide when completing the assessment.
 - a. Reinforce the need for the scoring guide to be used. The automated system should assist in the availability of the definitions for each item.
3. As counties incorporate the OYAS into other assessments, there have been times when questions did not get transferred to the modified interview guide.
 - a. Each county should formally review the interview guides used to ensure that all items are scored accurately.
4. Although within an acceptable range, the Pro-Social Skills domain demonstrated the lowest levels of interrater reliability across all three tools.
 - a. Special effort should be afforded to answer the questions in the Pro-Social Skills domain. This would include using follow-up questions more effectively, ensuring that adequate information is captured effectively.
 - b. Training should be modified to increase the focus on this section. Shorter vignettes could be incorporated into the training process to increase the opportunity to score these items.
 - c. OYAS scoring should be reviewed periodically during staff meetings to ensure interrater reliability is being maintained. These discussions can include focusing on the Pro-Social Skills domain.
 - d. UC will review the scoring guide to determine if there are areas that can be clarified to make scoring more accurately.

5. The training procedures adopted by Ohio should be maintained.
 - a. The training protocol has allowed for the OYAS to overcome traditional issues with scoring dynamic items resulting in acceptable ranges of interrater agreement.